

AN INTRODUCTION TO PROBABILITY DISTRIBUTIONS

ANTONY UGONI BSc (Hons).^{*}
BRUCE F. WALKER DC.[†]

Abstract: Probability allows us to infer from a sample to a population. In fact, inference is a tool of probability theory. This paper looks briefly at the Binomial, Poisson, and Normal distributions. These are probability distributions, which are used extensively in inference. An understanding of these distributions will assist the chiropractor and osteopath to critically appraise the literature.

PROBABILITY DISTRIBUTIONS

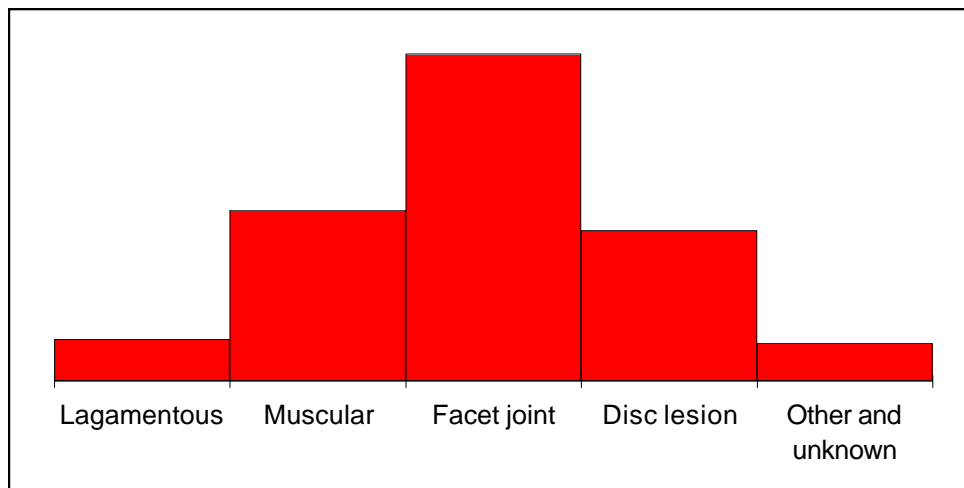
No matter how large or small a quantitative research study may be, it will always end up with a collection of values or numbers. When the various frequencies of these numbers are plotted, it will give a visual representation of their distribution. For example, in a survey of low back pain sufferers we might make the following diagnoses.

Key Indexing Terms: Probability distributions, biostatistics, chiropractic, osteopathy.

Diagnoses	Ligamentous	Muscular	Facet joint	Disc lesion	Other and unknown	Total
Numbers of cases	111	456	872	401	101	1941

The frequency, or probability distribution for the above example is shown below in figure 1

Figure 1



Before continuing, we will explain the terms *probability*, *independent*, and *mutually exclusive* events as this will assist in the understanding of the paper.

The probability of an event is the proportion of times an event will occur if the experiment is repeated infinitely. These probabilities lie between 0 and 1, and an event with probability 0 will never occur, while an event with probability 1 will always occur.

Most inferential techniques require independence of all events and observations. This means that the outcome of one event does not alter the probability on

^{*} DEPARTMENT OF SOCIAL AND PREVENTIVE MEDICINE,
MONASH UNIVERSITY. COMMERCIAL RD, PRAHRAN. VIC. 3181.
[†] PRIVATE PRACTICE OF CHIROPRACTIC.
33 WANTIRNA RD, RINGWOOD, VICTORIA, 3143. PH 03 879 5555

any other event. For example, tossing a coin and getting a head does not change the probability of a head on the next toss from being $1/2$. The probability of independent events occurring successively is found by multiplying the probabilities of these events. For example, the probability of tossing a coin twice and getting heads both time is $P(\text{Head}) \times P(\text{Head}) = (1/2) \times (1/2) = 1/4$.

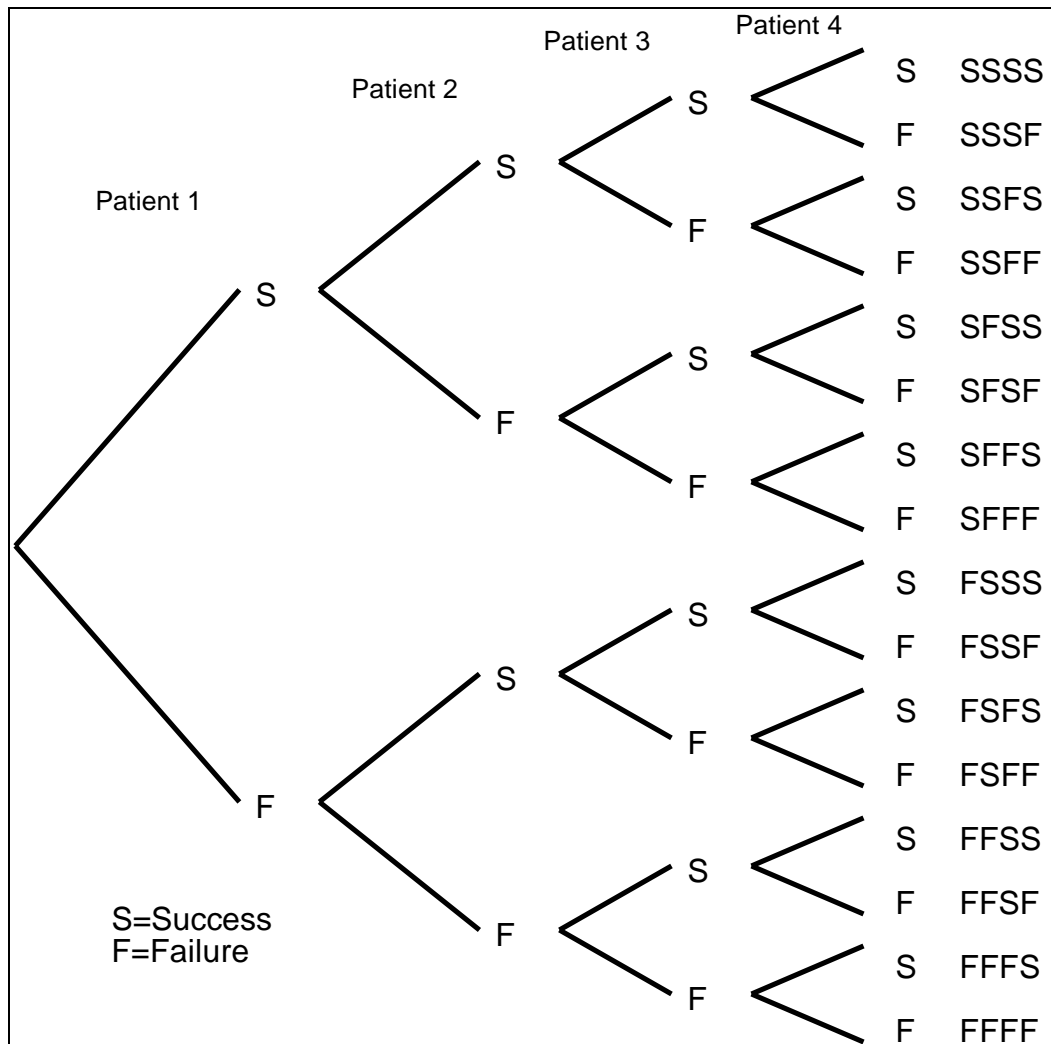
Two or more events are mutually exclusive if the occurrence of one event precludes the occurrence of others. For example, when tossing a coin, the toss of a head precludes a tail. The probability of an element of a group of mutually exclusive events occurring is found by adding the probabilities of the events. For example, on the role of a dice, the probability of a 5 or a 6 being thrown is $P(5) + P(6) = (1/6) + (1/6) = 2/3$

This paper examines the properties of three probability distributions used in research. An understanding of these distributions will assist in the critical appraisal of the literature and an appreciation of why certain statistical tests are selected to analyse the data.

THE BINOMIAL DISTRIBUTION

A chiropractor has 4 patients all presenting with benign occipital headache and she has been told by her colleagues that usually 3 out of 4 patients with this presentation will respond to neck manipulation with one treatment. She becomes interested in the probability of treatment being successful in 3 of the 4 patients. A 'Tree' of all possible events can be drawn out, and is shown in figure 2.

Figure 2



For example, if patients 1, 3 and 4 have successful treatment, and patient 2 is not, then this corresponds to following the 'branches' to the point SFSS (Success, Failure, Success, Success).

The original problem was to find the probability of 3 of 4 patients having successful treatments. This can happen in 4 ways, through the sequences SSSF, SSFS, SFSS, or FSSS. That is all but patient 4 having successful treatment, or all but patient 3 having successful treatment, etc.

To continue, we need to digress and consider the advantage of independence and identical probabilities of success.

Let 'H' denote a Head on the toss of a coin and 'T' denote Tails, and let $P(H)$ denote the probability that on any flip of the coin a head will result. What is the probability of rolling the sequence {Head, Tail, Head}? That is, what is $P(H \text{ and } T \text{ and } H)$? If each toss of the coin is independent of every other toss, and the probability of heads (or tails) remains identical at each toss, then we can write

$$P(H \text{ and } T \text{ and } H) = P(H)P(T)P(H) = P(H)^2P(T) = (\frac{1}{2})^2 \times \frac{1}{2} = 0.125$$

Armed with this knowledge, it is easy to show that in our example:

$$P(SSSF) = P(S)P(S)P(S)P(F) = P(S)^3P(F)$$

Similarly

$$\begin{aligned} P(SSFS) &= P(S)P(S)P(F)P(S) = P(S)^3P(F) \\ &= P(SFSS) \\ &= P(FSSS) \end{aligned}$$

Thus, the probability of seeing 3 successes out of 4 treatments (which can happen in 4 different ways) is

$$P(3 \text{ successes out of 4 treatments}) = 4P(S)^3P(F)$$

To simplify matters, if we consider the situation where only 1 of 2 possible outcomes for each patient are possible (Success and Failure) as opposed to multiple possible outcomes (eg Success, Partial success, Failure), we can rewrite $P(F)$ as $1-P(S)$. Thus

$$P(3 \text{ successes out of 4 treatments}) = 4P(S)^3(1-P(S))$$

If we follow the same logic, it is easy to show that for 2 successes out of 4, we get (SSFF, SFSF, SFSS, FSSF, FSFS, FFSS) ie 6 possible sequences. Therefore

$$P(2 \text{ successes out of 4 treatments}) = 6P(S)^2(1-P(S))^2$$

Following on with our example:

Let the probability of success be denoted as 'p', and let the actual probability of a successful treatment be 0.80 (ie $p=0.8$), we then have

$$\begin{aligned} P(3 \text{ successes out of 4 treatments}) &= 4P(S)^3(1-P(S)) \\ &= 4p^3(1-p) \\ &= 4(0.8)^3(1-0.8) \\ &= 4(0.512)(0.2) \\ &= 0.4096 \end{aligned}$$

and

$$\begin{aligned} P(2 \text{ successes out of 4 treatments}) &= 6P(S)^2(1-P(S))^2 \\ &= 6(0.8)^2(1-0.8)^2 \\ &= 6(0.64)(0.04) \\ &= 0.1536 \end{aligned}$$

In general, the probability of 'r' successes out of 'n' trials is equal to

$$\begin{aligned} P(r \text{ successes in } n \text{ trials}) &= \{ \text{number of sequences of} \\ &\quad n \text{ trials with } r \text{ successes} \} \\ &\quad \times \{ \text{probability of any such} \\ &\quad \text{sequence} \} \end{aligned}$$

The number (r) of successes in n trials is easy to calculate. It is simply $\frac{n!}{r!(n-r)!}$. The ! is called the factorial, and works in the following manner.

$$x! = x(x-1)(x-2)(x-3).....(x-x)!$$

For example,

$$\begin{aligned} 5! &= 5(5-1)(5-2)(5-3)(5-4)(5-5)! \\ &= 5(4)(3)(2)(1)(0)! \\ &= 5 \times 4 \times 3 \times 2 \times 1 \times 1 \\ &= 120. \end{aligned}$$

Notice that, $0!=1$, which is a mathematical convenience. To illustrate that $\frac{n!}{r!(n-r)!}$ yields the

desired result, recall that 3 successes out of 4 trials could happen 4 ways (SSSF, SSFS, SFSS, FSSS). In this example we have $n=4$, and $r=3$ and

$$\begin{aligned} \frac{n!}{r!(n-r)!} &= \frac{4!}{3!(4-3)!} = \frac{(4)(3)(2)(1)(0)!}{(3)(2)(1)(0)!(1)(0)!} = \frac{24}{(6)(1)} = 4 \\ &\text{as required.} \end{aligned}$$

Thus, the probability of r successes in n trials, each with identical probability of success p, is

$$\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

For example, if we were interested in the probability of 3 successes out of 5 treated patients, each with probability of 0.75 of success, we would have

$$\begin{aligned} n &= 5 \\ r &= 3 \\ p &= 0.75 \end{aligned}$$

$$\begin{aligned} \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} &= \\ \frac{5!}{3!(5-3)!} (0.75)^3 (1-0.75)^{5-3} &= \\ \frac{5!}{3!(2)!} (0.75)^3 (0.25)^2 &= \\ \frac{120}{(6)(2)} (0.75)^3 (0.25)^2 &= 0.2637 \end{aligned}$$

The number of successes (out of 5 trials), can take several values (0, 1, 2, 3, 4, and 5), thus it is known as a variable. Common practice is to call this variable X, and use 'r' as the observed number of successes. Instead of writing P(r successes in n trials), we simply write P(X=r).

An important point to notice is the sum of all probabilities, shown in table 1. This table shows the probabilities of all possible number of heads (successes) in 6 throws (trials) of a coin, when the probability of a head is equal to 1/2.

r	P(X=r)
0	0.0156
1	0.0937
2	0.2344
3	0.3125
4	0.2344
5	0.0937
6	0.0156
1.0000	

Thus the sum of all probabilities equals 1. This makes intuitive sense, since the probability of 0 heads, or 1 head, or 2 heads,, or 6 heads out of 6 throws is the probability that at least one of all possible events will occur.

This is useful for calculating the probability of a large number of successes in a large number of trials.

For example, suppose the chiropractor mentioned above is interested in the probability of successful treatments in at least 10 of the next 100 patients, when the probability of a successful treatment is 0.80 (and each patient is independent of each other patient). The

chiropractor would need to calculate the probability of 10 successful treatments in the next 100 patients, and add this to the probability of 11 successful treatments in the next 100 patients, then add this to the probability of 12 successful treatments in the next 100 patients,, then add this to the probability of 100 successful treatments in the next 100 patients.

If we let X denote the random variable 'number of successful treatments in 100 patients', then we need to calculate

$$P(X \geq 10) = P(X=10) + P(X=11) + P(X=12) + \dots + P(X=99) + P(X=100)$$

This is a tiresome task indeed, involving the calculation of 91 probabilities, and adding them together. The calculation can be made easier by utilising the fact that the sum of all probabilities is equal to 1. Thus

$$\begin{aligned} P(X \geq 10) &= P(X=10) + P(X=11) + P(X=12) + \dots + P(X=99) + P(X=100) \\ &= 1 - P(X < 10) \\ &= 1 - P(X \leq 9) \\ &= 1 - [P(X=9) + P(X=8) + P(X=7) + \dots + P(X=1) + P(X=0)] \\ &= 1 - 0.632 \\ &= 0.9368 \end{aligned}$$

A graph of the binomial distribution can be plotted also. Below are 3 graphs with n=10 trials, and different probabilities (see figures 3, 4, and 5).

Figure 3

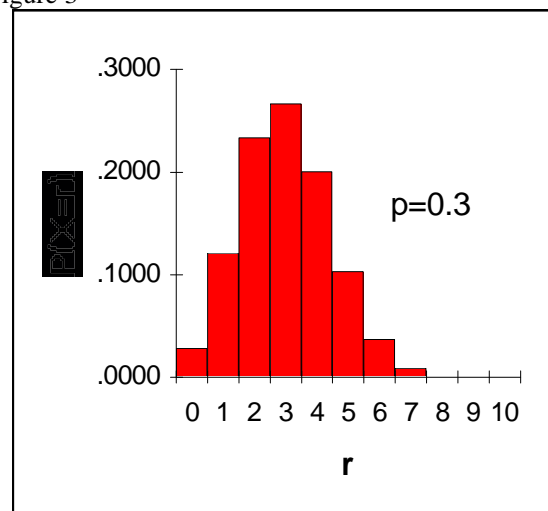


Figure 4

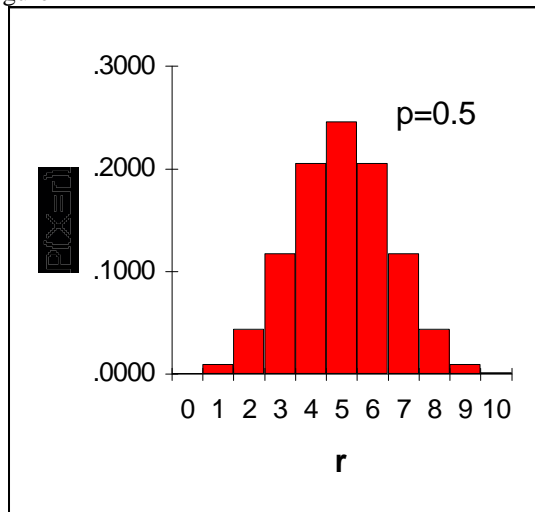
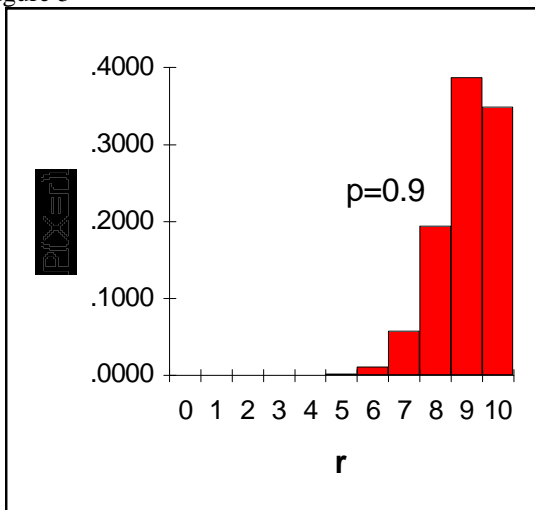


Figure 5



The horizontal axis represents each possible number of successes in 10 trials (treatments), and the vertical axis represents the probability of r successes in 10 trials. Important to note, is that the total area beneath each graph is equal to 1.

Notice that, with small probability of success (Figure 3), the graph is bunched (skewed) to the left. This means that most of the time, a small number of success will take place when the probability of success is small. Figure 5 is skewed to the right, indicating that a large number of successes will usually take place when the probability of success is high. Figure 4 shows an even spread of successes, which is expected since the probability of success is 0.5, or even chance.

The mean and standard deviation of the random variable X can easily be calculated with n (the number of trials) and p (the probability of success). These are

Mean = np

Standard Deviation = $\sqrt{np(1-p)}$

For example, with $n=10$ and $p=0.3$, we have

Mean = $10(0.3) = 3$

Standard Deviation = $\sqrt{10(0.3)(0.7)} = 1.45$

To summarise, the binomial distribution is useful for answering questions about the probability of X number of occurrences in n independent trials where there is a constant probability ' p ' of success in each trial.

THE POISSON DISTRIBUTION

Suppose we were interested in the number of rare but serious adverse effects (denoted by X) of chiropractic neck manipulation during the 1995 calendar year in Victoria.

What would be the probability that 2 such events would occur? That is, what is $P(X=2)$?

To calculate this probability, we must first know the average rate of adverse events in preceding years. This average rate (per unit time) is denoted by λ (lambda).

As it turns out, $\lambda=2.3$ adverse events per year.

Probabilities can be calculated using the following formula (the derivation of which is too involved for this paper).

$$P(X = r) = \frac{e^{-1} 1^r}{r!}$$

Where e is the exponential constant ($e=2.71828, \dots$).

We wish to know what the probability of only two adverse effects from chiropractic neck manipulation (when the average rate is 2.3) is for 1995. This is shown below.

$$P(X = 1) = \frac{e^{-2.3} 2.3^2}{2!} = 0.2652$$

That is, there is a 26.52% chance of two adverse effects in 1995.

Unlike the Binomial distribution, the probabilities for all events cannot be tabulated. This is due to the (theoretical) possibility of an infinite number of events (adverse effects) occurring in the unit time (year). that is, we can legitimately ask 'What is the probability of 100 (or 200, or 1000) adverse effects occurring in one year?'. Realistically, we would not expect more than 7 events when the average rate is 2.3 (probability of more than 7 events is 0.0033). The probabilities for adverse effects is tabulated below (up to 4 decimal places).

r	P(X=r)
0	0.1003
1	0.2306
2	0.2652
3	0.2033
4	0.1169
5	0.0538
6	0.0206
7	0.0068
8	0.0019
9	0.0005
10	0.0001
11	0.0000
	1.0000

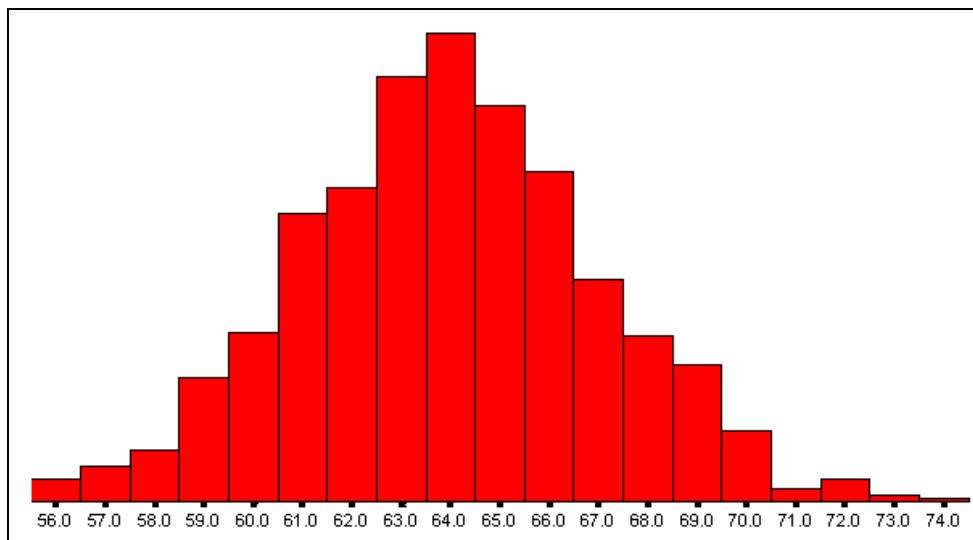
This distribution function is known as the Poisson distribution and requires only that each event is independent and has the same average rate.

THE NORMAL DISTRIBUTION

Often in medical research, data is continuous. Such examples are heart rates (eg 62.4 beats/minute), weight (eg 67.45 kg), and degrees of angulation in scoliosis. These values can (theoretically) take on an infinite number of values, and have no limit as to the accuracy with which they can be measured.

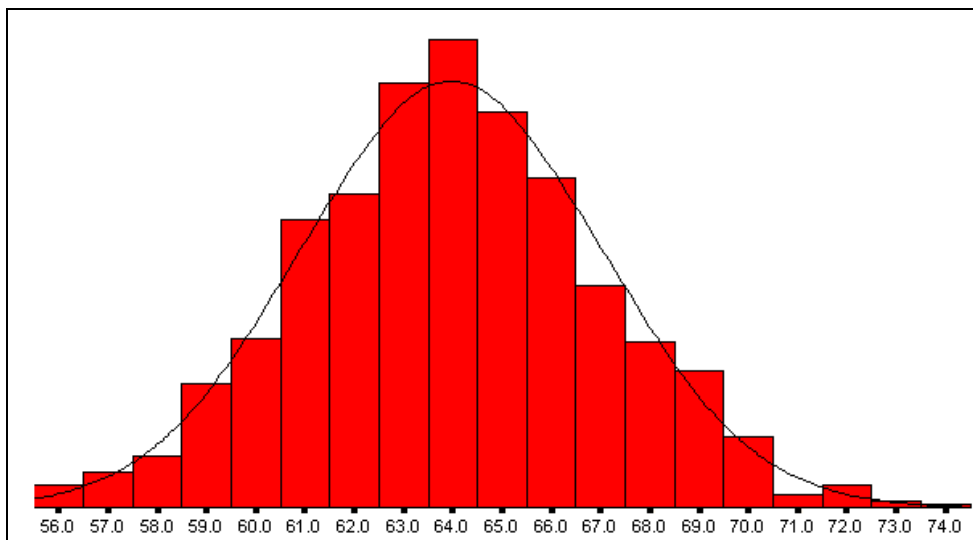
For example, suppose 1000 males (age 25) had their heart rates measured. taking all 1000 heart rates, we can construct the following relative frequency histogram (shown in figure 6).

Figure 6



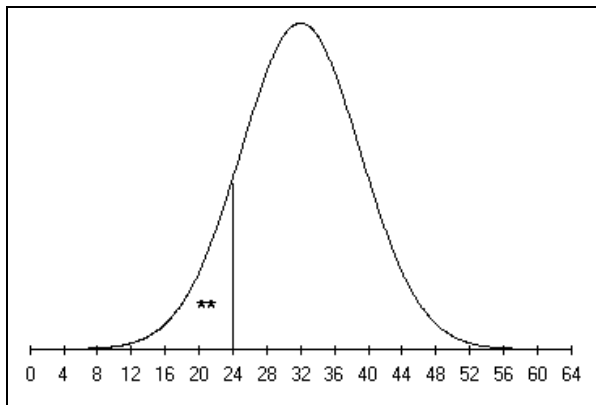
These histograms are difficult to work with, and so a common strategy is to use some form of a smoothed histogram such as the one below in figure 7

Figure 7



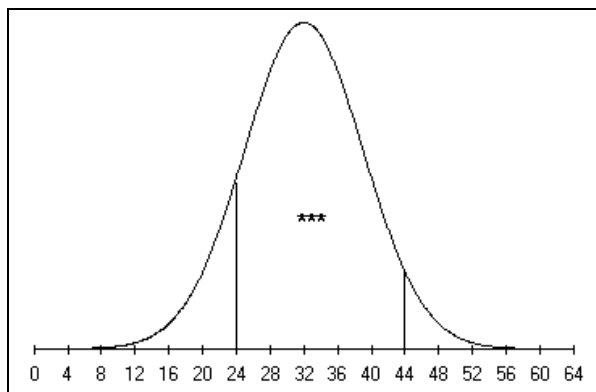
A feature of these curves is that the entire area below the curve can be scaled to equal 1, and partial areas correspond to probabilities. For example, if we let 'A' be the Cobb angle of a randomly chosen scoliotic female, then the probability of her Cobb angle being less than 24° can be written as $P(A < 24)$. This probability corresponds to the area (marked **) to the left of the point 24 (shown in figure 8).

Figure 8



Similarly, the probability that a randomly chosen female will have a Cobb angle between 24° and 44°, $P(24 < A < 44)$, is the area within the interval (***) confined by 24 and 44 (shown in figure 9).

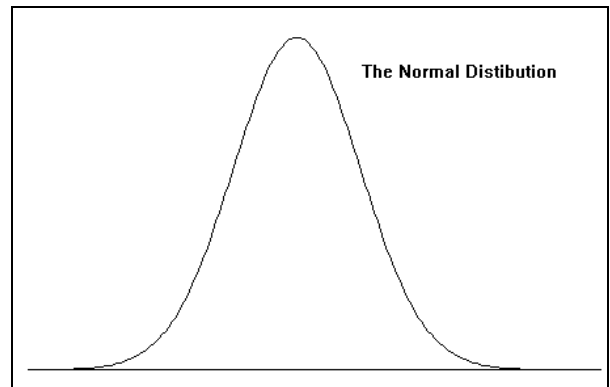
Figure 9



In general, the use of the smoothed curve is to show how a population may vary.

The most common of these curves is that of the Normal Distribution. The general shape of this curve is commonly called 'bell shaped' and has the form shown in figure 10.

Figure 10



The mathematical form of the curve is

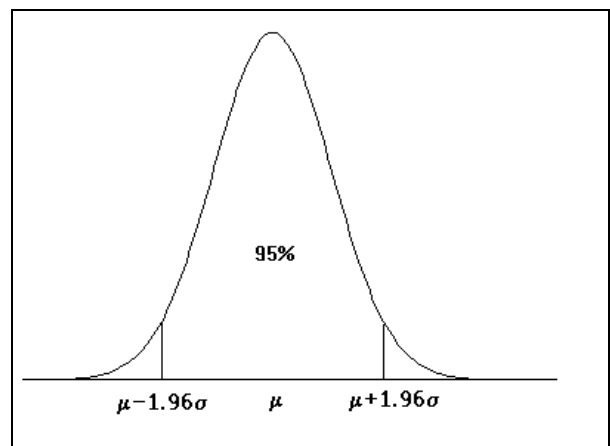
$$f(x) = \frac{1}{\sqrt{2\pi}s} \exp\left\{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right\}$$

and is characterised by the mean, μ , and the standard deviation σ .

The normal curve has the following properties.

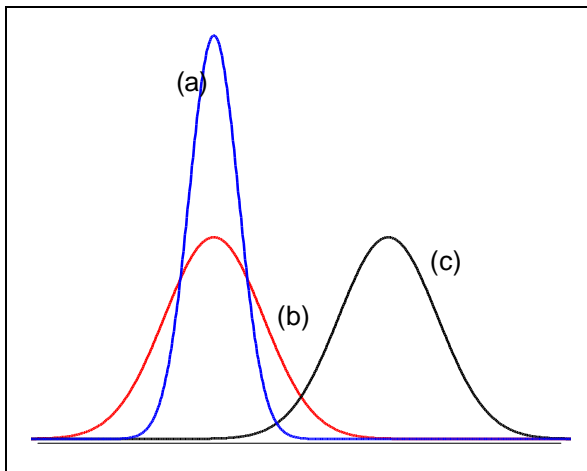
1. Symmetric about the mean μ
2. The curve gets closer and closer to the horizontal axis but never touches it.
3. the total area beneath the curve is 1.
4. 68% of the area lies between $\mu - \sigma$, and $\mu + \sigma$ ie 1 standard deviation, 95% of the area lies between $\mu - 1.96\sigma$, and $\mu + 1.96\sigma$ (shown below in figure 11) ie 2 standard deviations, 99.75% of the area lies between $\mu - 3\sigma$, and $\mu + 3\sigma$ ie 3 standard deviations.

Figure 11



The mean and standard deviation affect the curve in the following ways

Figure 12



The standard deviation affects the spread of the data. This is shown with curves (a) and (b) (figure 12), which have the same mean but different standard deviations. The mean affects the location of the data. This is shown with curves (a) and (c), which have the same standard deviations, but different means.

The application of such a curve will be displayed using the following example:

The heights of 16 year old girls have a normal distribution with mean $\mu=140\text{cm}$ and standard deviation $\sigma=5\text{cm}$. Let 'X' be the random variable 'Height' for a 16 year old girl.

What is

- 1) The probability that a randomly chosen girl's height is between 135cm and 145cm?
- 2) The probability that a randomly chosen girl's height is between 130.2cm and 149.8cm?
- 3) The probability that a randomly chosen girl's height is between 130cm and 150cm?

These are calculated below:

- 1) $P(135 < X < 145) = P(140 - 5 < X < 140 + 5)$
 $= P(\mu - \sigma < X < \mu + \sigma)$
 $= 0.68$
- 2) $P(130.2 < X < 149.8) = P(140 - 1.96 \times 5 < X < 140 + 1.96 \times 5)$
 $= P(\mu - 1.96\sigma < X < \mu + 1.96\sigma)$
 $= 0.95$
- 3) $P(130 < X < 150) = P(140 - 2 \times 5 < X < 140 + 2 \times 5)$
 $= P(\mu - 2\sigma < X < \mu + 2\sigma)$
 $= 0.9975$

By now, the reader might have observed that probabilities from a normal distribution are dependent solely on the number of standard deviations an observation lies from its mean. Thus, to calculate the probability that a 16 year old girl is more than 157cm (mean $\mu=140\text{cm}$ and standard deviation $\sigma=5\text{cm}$), we calculate, using the following formula (trust us)

$$\frac{\text{observation} - \text{mean}}{\text{standard deviation}}$$

For the probability of a 16 year old being more than 157cm, we have

$$\frac{157 - 140}{5} = \frac{17}{5} = 3.4$$

This calculation is known as the Z-score, and tables with corresponding probabilities exist for these scores in most statistics texts.

DISCUSSION

We have discussed three common probability distributions; Binomial, Poisson, and Normal.

The Binomial distribution is used to model events that have a dichotomous outcome (eg. success or failure), and to determine the probability of outcomes of interest.

The Poisson distribution is used to determine probabilities where (possibly unlimited) counts are of interest eg. Serious adverse affects in neck manipulation.

The Normal distribution is used to determine probabilities of data on a continuous numerical nature. These distributions are often used in clinical studies. it is important to recognise them and therefore assist with a critical appraisal of the literature.

RECOMMENDED READING

1. Altman D. Practical Statistics for Medical Research. Chapman and Hall, London, 1992
2. Spatz C., Johnston J. O. Basic Statistics: Tales of Distributions. Brooks/Cole, 1989
3. Dawson-Saunders B., Trapp R.G. Basic and Clinical Biostatistics. Lange: Prentice Hall, Connecticut, 1990